
Chapter 5: Item Nonresponse

This section examines and quantifies the extent of missing data, formally called item nonresponse, in NLSY79 surveys. To provide readers with a detailed view of this problem, four surveys are analyzed. Nonresponse rates are examined first in the 1979 survey and then in the surveys that occur at five year intervals (1984, 1989, and 1994). These years were chosen to capture the major changes in the NLSY79 survey. Examining the 1979 survey shows the initial levels of nonresponse. Examining the 1984 survey shows the amount of nonresponse in the survey just before one part of the respondent pool was dropped for funding reasons. The 1989 data show nonresponse after the first set of NLSY79 respondents were dropped. The 1994 data are representative of what occurs after users and interviewers are switched from paper and pencil interviewing (PAPI) to computer assisted personal interviewing (CAPI).

This section focuses on the three types of missing data: refusals, invalid skips, and don't knows. Overall, the section shows that in these four NLSY79 surveys, 12.5 million questions were asked. Out of all the questions asked to respondents, about 1.6 percent do not have valid answers and are missing data. Of the three missing data categories, the majority of the missing data, comprising one percent of all questions, are invalid skips.

5.1 Introduction

Missing data, or nonresponse, happens in a number of ways in the NLSY79 survey. First, a number of respondents do not participate at all, causing all information in that particular survey to be missing. The extent of non-participation in each survey round is quantified in Chapter 2 and the particular reason by the created variable labeled "Reason for Noninterview." Readers interested in understanding how many individuals refuse to participate should look at this chapter.

A second reason missing data occurs is that respondents do not provide a valid answer to a question. When this happens, interviewers make a determination about whether to mark the answer as a refusal or don't know value. Users should be cautioned that the assignment of refusals and don't knows is likely to vary across interviewers. Moreover, some respondents may believe it is impolite to refuse a question and decline to answer by saying they do not know. Hence, whether a question is marked either a refusal or a don't know is somewhat arbitrary.

The last major way missing data occurs is when the interviewer incorrectly follows the survey instrument's flow. Incorrect flows result in some respondents being skipped over a set of questions that should be answered while others answer questions that they should not have been asked. NLS data archivists have removed from the data most of the extraneous question responses. While extra information can be removed, missing data is not imputed in the NLSY79 surveys. Missing data caused by this reason is flagged with a special "invalid skip" code. Readers should note that the number of

invalid skips drops precipitously beginning in 1993 with the introduction of CAPI. Nevertheless, invalid skips are still possible in CAPI data. If the CAPI survey contains a programming mistake, the instrument could incorrectly sequence a respondent. When these errors are found, the CAPI survey is patched in the field to prevent further invalid skips but the incorrect cases are not refiled.

All missing data are clearly flagged in the NLSY79 data set. Five negative numbers are used to indicate to user that the variable does not contain useful information. The five values are: (-1) refusal, (-2) don't know, (-3) invalid skip, (-4) valid skip, and (-5) noninterview. These five numbers are reserved as missing value flags and, with a few exceptions (see appendix 5 in the *NLSY79 Codebook Supplement*), are rarely used in the NLSY79 for valid data values.

In the tables that follow, only variables in a given survey year that were filled in by either a respondent or an interviewer are analyzed. Thus all created, machine check, date and time stamp, and variables generated in data post-processing are excluded from the analysis.

This section is not the only research on the extent of missing data in the NLS. Olsen (1992) investigated the effect of switching from PAPI to CAPI interviewing. His research shows fewer interviewer errors occur from navigating the instrument as well as fewer don't knows in the CAPI survey. More importantly, CAPI respondents appeared more willing to reveal sensitive material in the alcohol use section. Mott (1985, 1984, and 1983) examines the NLSY79's fertility data. In these reports, he examines the 1982 and 1983 surveys and finds very low refusal rates for the data in general. However, by shifting to a confidential abortion reporting method, the willingness to respond greatly increases. Mott (1998) examines the amount of missing data about the children of NLSY79 females. He finds that Hispanics and, to a smaller extent blacks, have a much higher probability of not finishing the child assessments after starting the interview.

The rest of this chapter contains three parts. The next part examines which sections of the NLSY79 have high nonresponse rates. Then, responses are examined to see how many times individuals do not respond to questions. The last section examines which particular questions in sections with high nonresponse rates are causing problems.

5.2 Item Nonresponse by Section

Which parts of the NLSY79 survey have the highest rates of nonresponse? This section examines the 1979, 1984, 1989, and 1994 surveys and shows which portions have the most missing data. The extent of nonresponse is shown for each year by the tables that examine every survey section. The first column of the tables contains the section names within the survey. The second column shows the total number of questions that all respondents and all interviewers should have answered in that section. This

number is determined by first calculating within each section the number of questions each respondent should answer. A question is considered answerable if it does not have a valid skip (-4) or non-interview (-5) as its answer. A total for the section is obtained by summing up the answers for all NLSY79 respondents.

The third (don't know), fourth (refusal), and fifth (invalid skip) columns show the total number of nonresponses found in each section. Columns six, seven, and eight show the same information except in percentage form. The ninth column shows the total percentage of questions missed and is the sum of the previous three percentages. The last column, labeled rank, shows which sections have the most (closer to 1) and least (further from 1) amount of nonresponse.

The bottom row of each table combines the information and shows totals. For example, the bottom of the "Number Questions Asked" column in the 1979 survey shows that almost four million questions (3,975,146) were expected to be filled in by respondents or interviewers. While the 1979 survey contained the most questions, the other years are not far behind. In 1984, there were 3,067,473 questions, while in 1989 and 1994 there were 1,793,721 and 3,669,260 questions respectively. Readers are cautioned that each year of NLSY79 data contains far more variables since the tables exclude questions labeled machine checks, date and time stamp, and questions with valid skip or noninterview data flags.

The four tables show that the overall rate of missing data has dropped steadily over time. In 1979, 2.7 percent of the questions in the survey were not answered. This number drops to 1.9 percent in 1984 and then falls to 0.9 percent in 1989 and 0.7 percent in 1994. Hence, nonresponse problems are less of a concern in later rounds since it appears the most willing respondents continue to participate in the survey.

Combining the data from all sections in all four tables shows the majority of nonresponse is caused by invalid skips. The four surveys asked a total of 12.5 million questions. Of these questions, almost 125,000, or one percent, were invalid skips; 69,000, or half of one percent, were don't knows; and only 11,000 or 0.08 percent were refusals. This shows that in these four surveys the primary nonresponse problem is invalid skips, not refusals or don't knows.

Examining the four tables over time shows a steady decrease in the amount of data missing due to invalid skips. In 1979, invalid skips comprised 2.1 percent of the questions asked. This number dropped sharply to 1.2 percent by 1984 and then down to 0.25 percent by 1989. The switch to CAPI almost completely eliminates invalid skips; only 57 questions out of almost 3.7 million were incorrectly skipped.

While invalid skips fall over time, the percentage of refusals increases and the percentage of don't knows stays steady. Don't knows comprise slightly more than half of a percent of all questions asked in each of the four surveys. Specifically, don't knows comprised 0.6 percent in 1979, 0.6 percent in 1984, 0.5 percent in 1989, and 0.5 percent in 1994. Refusals steadily increase over time and comprise 0.01 percent in 1979, 0.07 percent in 1984, 0.10 percent in 1989, and 0.16 percent in 1994.

The last column, labeled rank, shows that missing data are not confined to a single section or area of the survey. Table 5.2.1 shows that in 1979 the work experience section, with 14.5 percent of the questions missing valid data, had the most problems. Fourteen percent of all questions asked in this section are labeled as invalid skips and only 0.5 percent of the questions were either refusals or don't knows. Military experience, the second most problematic section had almost half the rate of missing data (7.8 percent) as work experience. The table shows the problem of invalid skips is not related to subject matter since the section (rank 21 out of 21) with least problems, titled "on jobs", also focuses on labor market issues, like work experience.

Table 5.2.1 Extent of Refusals, Don't Knows & Invalid Skips in 1979 NLSY79 Survey

Section Name	# Questions Asked	# Don't Knows	# Refused	# Invalid Skipped	% Don't Knows	% Refused	% Invalid Skipped	Total % Missed	Rank
Family Background	660803	6196	90	12292	0.94%	0.01%	1.86%	2.81%	7
Marital Status	32995	131	25	467	0.40%	0.08%	1.42%	1.89%	14
Fertility	82141	679	23	624	0.83%	0.03%	0.76%	1.61%	17
Schooling	402134	994	14	5592	0.25%	0.00%	1.39%	1.64%	16
Pay	211504	22	0	3482	0.01%	0.00%	1.65%	1.66%	15
World of Work	220185	2220	31	2883	1.01%	0.01%	1.31%	2.33%	10
Military	145619	491	24	10885	0.34%	0.02%	7.47%	7.83%	2
CPS	396697	862	8	10969	0.22%	0.00%	2.77%	2.98%	5
On Jobs	230982	135	2	903	0.06%	0.00%	0.39%	0.45%	21
Employer Supp.	291836	2009	69	3575	0.69%	0.02%	1.23%	1.94%	13
Last Job	44504	31	0	261	0.07%	0.00%	0.59%	0.66%	20
Work Experience	67695	288	15	9476	0.43%	0.02%	14.00%	14.45%	1
Gov. Training	36728	62	28	2124	0.17%	0.08%	5.78%	6.03%	3
Other Training	103662	52	0	2936	0.05%	0.00%	2.83%	2.88%	6
Not at Work	90768	79	7	5019	0.09%	0.01%	5.53%	5.62%	4
Health	67869	358	2	545	0.53%	0.00%	0.80%	1.33%	18
Significant Others	58816	669	0	585	1.14%	0.00%	0.99%	2.13%	12
Residences	52845	94	7	1029	0.18%	0.01%	1.95%	2.14%	11
Rotter Scale	202976	1277	15	521	0.63%	0.01%	0.26%	0.89%	19
Income & Assets	321685	1667	216	6813	0.52%	0.07%	2.12%	2.70%	8
Expectations	252702	3824	20	2092	1.51%	0.01%	0.83%	2.35%	9
Total	3975146	22140	596	83073	0.56%	0.01%	2.09%	2.66%	–

While the “on jobs” section of the survey consistently has the least problems in these four surveys, the section with the most problems changes. Table 5.2.2, which examines the 1984 survey, shows the most problems in the “Fertility” section. Of the almost half-million questions asked in the fertility section, 5.6 percent contain missing data. While the majority of problems (3.4 percent) were due to invalid skips, a surprisingly large 2 percent of the missing responses are don’t knows. The second most problematic section in the 1984 survey was “Drug Use” where 2.7 percent of the questions have missing data. Like “Fertility”, the major portion of the problem is invalid skips (1.8 percent), but don’t knows (0.8 percent) also comprise a significant share. Interestingly, refusals comprise only 0.1 percent, a relatively small proportion for a sensitive topic, suggesting that some of the don’t knows were hidden refusals.

Table 5.2.2 Extent of Refusals, Don’t Knows & Invalid Skips in 1984 NLSY79 Survey

Section Name	# Questions Asked	# Don’t Knows	# Refused	# Invalid Skipped	% Don’t Knows	% Refused	% Invalid Skipped	Total % Missed	Rank
Calendar	88462	8	0	4	0.01%	0.00%	0.00%	0.01%	15
Marital Status	50206	273	18	561	0.54%	0.04%	1.12%	1.70%	4
Schooling	324139	1031	469	2164	0.32%	0.14%	0.67%	1.13%	9
Military	123126	337	41	1352	0.27%	0.03%	1.10%	1.41%	7
CPS	333267	467	5	4270	0.14%	0.00%	1.28%	1.42%	6
On Jobs	140382	0	0	17	0.00%	0.00%	0.01%	0.01%	16
Gaps in Jobs	120601	15	0	175	0.01%	0.00%	0.15%	0.16%	13
Gov. Training	31226	38	0	59	0.12%	0.00%	0.19%	0.31%	12
Other Training	45002	7	0	736	0.02%	0.00%	1.64%	1.65%	5
Fertility	462288	9141	891	15739	1.98%	0.19%	3.40%	5.57%	1
Child Care	114317	201	13	1157	0.18%	0.01%	1.01%	1.20%	8
Health	52866	35	3	29	0.07%	0.01%	0.05%	0.13%	14
Alcohol	314511	33	47	2234	0.01%	0.01%	0.71%	0.74%	11
Drug Use	414007	3464	300	7454	0.84%	0.07%	1.80%	2.71%	2
Income & Assets	439646	2945	241	938	0.67%	0.05%	0.21%	0.94%	10
Attitudes	13427	214	2	29	1.59%	0.01%	0.22%	1.82%	3
Total	3067473	18209	2030	36918	0.59%	0.07%	1.20%	1.86%	–

Table 5.2.3 shows the amount of nonresponse in the 1989 survey. The most problematic section is income, missing data in 1.3 percent of its questions, with the CPS’s (Current Population Survey) 1.2 percent rate in a close second. Unlike earlier years, the major missing data problem in both the income (1 percent) and CPS (0.8 percent) sections are don’t knows, not invalid skips (0.1 percent income and 0.4 percent CPS).

Table 5.2.3 Extent of Refusals, Don't Knows & Invalid Skips in 1989 NLSY79 Survey

Section Name	# Questions Asked	# Don't Knows	# Refused	# Invalid Skipped	% Don't Knows	% Refused	% Invalid Skipped	Total % Missed	Rank
Intro.	14647	20	1	41	0.14%	0.01%	0.28%	0.42%	7
Marital	86563	372	121	450	0.43%	0.14%	0.52%	1.09%	3
Schooling	76999	179	39	217	0.23%	0.05%	0.28%	0.56%	6
Military	33579	1	1	40	0.00%	0.00%	0.12%	0.13%	10
CPS	406265	3320	52	1650	0.82%	0.01%	0.41%	1.24%	2
On Jobs	39749	0	0	1	0.00%	0.00%	0.00%	0.00%	12
Gaps	91565	91	1	894	0.10%	0.00%	0.98%	1.08%	4
Gov. Training	49657	118	35	233	0.24%	0.07%	0.47%	0.78%	5
Fertility	152546	6	35	92	0.00%	0.02%	0.06%	0.09%	11
Health	154024	120	74	168	0.08%	0.05%	0.11%	0.24%	9
Alcohol	217441	74	400	201	0.03%	0.18%	0.09%	0.31%	8
Income	470686	4761	1124	439	1.01%	0.24%	0.09%	1.34%	1
Total	1793721	9062	1883	4426	0.51%	0.10%	0.25%	0.86%	–

Table 5.2.4 shows that the most problematic area in the 1994 survey includes the asset questions, which are missing 2.5 percent of their answers (75 percent of those missing being don't knows). The second most problematic area includes income questions, which are missing 1.3 percent of their answers. While in the three previous surveys refusal rates were not an issue, the 1994 survey shows refusals are becoming significant. Slightly more than half a percent (0.6 percent) of the asset section questions and more than one fifth of a percent (0.2 percent) of the income section questions were refused.

Table 5.2.4 Extent of Refusals, Don't Knows & Invalid Skips in 1994 NLSY79 Survey

Section Name	# Questions Asked	# Don't Knows	# Refused	# Invalid Skipped	% Don't Knows	% Refused	% Invalid Skipped	Total % Missed	Rank
Intro.	36251	62	14	0	0.17%	0.04%	0.00%	0.21%	12
Marital Status	137540	1522	193	0	1.11%	0.14%	0.00%	1.25%	3
School	60166	302	2	0	0.50%	0.00%	0.00%	0.51%	7
Military	27372	6	1	0	0.02%	0.00%	0.00%	0.03%	15
CPS	269452	28	9	0	0.01%	0.00%	0.00%	0.01%	17
On Jobs	79567	6	7	0	0.01%	0.01%	0.00%	0.02%	16
Employ. Suppl.	1060679	7092	1342	8	0.67%	0.13%	0.00%	0.80%	5
Training	194147	246	29	47	0.13%	0.01%	0.02%	0.17%	13
Fertility	450871	1859	763	0	0.41%	0.17%	0.00%	0.58%	6
Child Care	26453	109	12	0	0.41%	0.05%	0.00%	0.46%	9
Relationship	81477	285	113	0	0.35%	0.14%	0.00%	0.49%	8
Health	282702	623	199	0	0.22%	0.07%	0.00%	0.29%	11
Alcohol	164663	46	61	0	0.03%	0.04%	0.00%	0.06%	14
Income	305693	3176	672	1	1.04%	0.22%	0.00%	1.26%	2
Prog. Participation	118305	297	63	0	0.25%	0.05%	0.00%	0.30%	10
Assets	169301	3239	930	1	1.91%	0.55%	0.00%	2.46%	1
Drugs	204621	772	1626	0	0.38%	0.79%	0.00%	1.17%	4
Total	3669260	19670	6036	57	0.54%	0.16%	0.00%	0.70%	–

5.3 Item Nonresponse by Respondents

This section provides details on the amount of missing data associated with each respondent. Each table in this section shows the number of respondents who are missing data in one of the four surveys. The tables are split into two parts. The left hand part, columns one to four, shows the total number of questions that have missing data for each group of respondents. The right hand part, columns five to nine, shows the percentage of questions that have missing data.

The top line of table 5.3.1 shows that in the 1979 survey, 12,527 respondents never refused to answer questions. While refusals are quite rare in this survey round, don't knows and incorrect skips are quite frequent. The top line shows that only 5,084 respondents had zero don't know responses and only 2,347 respondents were sent through the entire questionnaire without any sequencing errors. Subtracting these numbers from the 12,686 total respondents means that 60 percent, or 7,602 respondents, stated they did not know the answer to at least one question and 81.5 percent, or 10,339 respondents, were incorrectly skipped somewhere in that questionnaire.

The right hand side of Table 5.3.1, which examines the percentage of questions missing data, shows a similar picture. Refusal rates are relatively low. There are 12,620 respondents who refused less than one percent of their questions, which means only 66 respondents refused one percent or more of the questions they were expected to answer. Thirty-five percent, or 8,185 respondents, answered don't know to less than one percent of their questions. Again, the largest group were respondents who were incorrectly skipped over questions. Only 4,313 respondents were incorrectly skipped over less than one percent of the questions, but 8,373 of the respondents were illegally skipped over one percent or more of their questions and 227 were skipped over more than 10 percent.

Refusal rates have increased steadily over time even though the more difficult respondents have presumably left the survey. Table 5.3.2, which examines the 1984 survey, shows an increase over the 1979 refusal rates. While the number of respondents answering the survey is shrinking, the number refusing to answer questions is increasing. For example, while in 1979 only 10 respondents refused to answer more than 10 questions, in 1984 there were 41 respondents. This pattern of increase is evident in Table 5.3.3, which examines 1989, and Table 5.3.4, which examines 1994. By 1994, there were 138 respondents who refused to answer more than 10 questions. Increasing refusal rates are also seen in the percentage side of the table. In 1979, only 43 respondents refused to answer one percent of the questions they were asked. This increased in subsequent surveys to 207 respondents in 1984, 193 respondents in 1989, and 246 respondents in 1994.

Don't know rates have also risen over time. In the 1979 survey, 8,185 respondents had less than one percent of their questions labeled as don't knows. This number drops in 1984 to 7,003 respondents and

further drops to 6,423 in 1989 and 5,942 in 1994. While rates have risen, relatively few individuals have high levels of don't knows. In 1979, only 68 respondents didn't know the answer to more than five percent of the questions they were asked. This number falls to 19 respondents in 1984 and then rises to 66 in 1989 before falling back to 46 respondents in 1994.

While don't know and refusal rates have risen, incorrect skip problems have clearly shrunk over time. In 1979, there were only 2,347 respondents who were correctly sequenced through the entire survey. In 1984, this number rises to 7,802 respondents, followed by a rise to 9,334 respondents in 1989. In 1994, almost every respondent was correctly sequenced. Only 57 respondents were incorrectly skipped through part of the survey and each respondent was only incorrectly skipped in a single question.

Table 5.3.1 Number of Respondents with Missing Data in 1979 Survey

Number of Questions	Number of Respondents			Percent of Questions	Number of Respondents		
	Refused	Didn't Know	Was Incorrectly Skipped Over		Refused	Didn't Know	Was Incorrectly Skipped Over
0	12527	5084	2347	0%	12620	8185	4313
1	91	2974	1897	1%	43	3247	3421
2	26	1723	1393	2%	7	773	1733
3	13	1016	1158	3%	5	264	989
4	5	629	838	4%	5	101	621
5	2	376	596	5%	0	48	397
6	1	228	489	6%	2	27	312
7	3	173	502	7%	1	18	278
8	3	131	420	8%	1	6	206
9	1	84	340	9%	0	7	118
10	4	57	308	10%	0	2	71
> 10	10	211	2398	> 10%	2	8	227

Table 5.3.2 Number of Respondents with Missing Data in 1984 Survey

Number of Questions	Number of Respondents			Percent of Questions	Number of Respondents		
	Refused	Didn't Know	Was Incorrectly Skipped Over		Refused	Didn't Know	Was Incorrectly Skipped Over
0	11222	4549	7802	0%	11749	7003	8956
1	610	3012	1289	1%	207	3807	1267
2	73	1901	622	2%	44	944	674
3	44	1136	413	3%	13	213	284
4	38	668	252	4%	15	62	133
5	13	345	369	5%	13	21	84
6	6	177	174	6%	10	11	139
7	1	108	93	7%	4	2	137
8	7	63	115	8%	5	3	107
9	4	38	73	9%	2	0	68
10	10	28	64	10%	2	3	36
> 10	41	44	803	> 10%	5	0	184

Note: Not included in this table are 617 respondents who did not answer the survey.

Table 5.3.3 Number of Respondents with Missing Data in 1989 Survey

Number of Questions	Number of Respondents			Percent of Questions	Number of Respondents		
	Refused	Didn't Know	Was Incorrectly Skipped Over		Refused	Didn't Know	Was Incorrectly Skipped Over
0	10221	6135	9334	0%	10250	6423	9461
1	171	2517	781	1%	193	3221	843
2	59	1036	189	2%	58	561	51
3	37	395	35	3%	35	219	69
4	20	194	20	4%	13	76	86
5	21	131	16	5%	10	39	24
6	7	75	7	6%	4	24	10
7	10	34	125	7%	4	17	10
8	10	24	18	8%	3	1	5
9	4	10	9	9%	3	3	9
10	7	6	3	10%	3	8	3
> 10	38	48	68	> 10%	29	13	34

Note: Not included in this table are 2,081 respondents who did not answer the survey.

Table 5.3.4 Number of Respondents with Missing Data in 1994 Survey

Number of Questions	Number of Respondents			Percent of Questions	Number of Respondents		
	Refused	Didn't Know	Was Incorrectly Skipped Over		Refused	Didn't Know	Was Incorrectly Skipped Over
0	7168	3559	8832	0%	8409	5942	8889
1	1129	1780	57	1%	246	2060	0
2	191	1082	0	2%	81	558	0
3	87	693	0	3%	41	165	0
4	41	443	0	4%	31	79	0
5	28	334	0	5%	20	39	0
6	29	232	0	6%	19	16	0
7	22	171	0	7%	6	15	0
8	21	115	0	8%	10	4	0
9	17	105	0	9%	9	2	0
10	18	72	0	10%	4	2	0
> 10	138	303	0	> 10%	13	7	0

Note: Not included in this table are 3,797 respondents who did not answer the survey.

5.4 Item Nonresponse within Problem Sections

How much missing data are associated with particular questions? This part of the chapter provides readers with an in-depth view of the questions within survey sections having a high amount of missing data. Like the previous parts, this section provides tables for four survey years. The first table (Table 5.4.1) examines questions from the 1979 survey's "Work Experience" section. This section has more missing data (14.5 percent) than any other 1979 survey section. The second set of tables (Tables 5.4.2 through 5.4.6) examines the most problematic section of the 1984 survey, "Fertility and Abortion." The third set of tables (Tables 5.4.7 and 5.4.8) examines the most problematic 1989 survey section, "Income and Assets." Since the 1994 "Income and Asset" section again ranked first in missing data, the last set of tables (Tables 5.4.9 and 5.4.10) substitutes the "Drug and Alcohol Use Supplements" given the high degree of research interest in understanding nonresponse in these sections.

To ensure the sets of tables are not overwhelming, all sections, like fertility, that could be naturally divided are split. Additionally, only the most important question or questions with high rates of nonresponse are shown.

Table 5.4.1, which examines the amount of missing data in the 1979 survey, shows the highest amount of missing data are associated with a pair of retrospective questions that asked respondents to remember what happened two years earlier. Interviewers incorrectly skipped slightly less than 1,750 respondents over R01150., weeks worked in 1977, and R01153., hours worked per week in 1977. Examining the 1979 questionnaire shows that these questions appear at the bottom of a page. Prior to these questions

is a fairly complicated half page of instructions and questions that the interviewer must read, understand, and partially speak. It seems likely that many interviewers did not understand the instructions and skipped to the next page.

Table 5.4.1 Amount of Missing Data Per Question in the Work Experience Section, 1979 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R01150.	Weeks Work in 1977	1735	11	1
R01151.	Weeks Work in 1976	418	18	1
R01152.	Weeks Work in 1975	240	11	0
R01153.	Hours/Week Work in 1977	1749	13	0
R01154.	Hours/Week Work in 1976	459	16	0
R01165.	Industry of 1 st Job after School	628	4	1
R01166.	Occupation at 1 st Job after School	627	3	1
R01167.	Hours/Week Work at 1 st Job after School	631	6	1
R01168.	Hours/Day at 1 st Job after School	632	6	1
R01169.	Rate of Pay at 1 st Job after School	632	32	2

Tables 5.4.2–5.4.6, which examine the “Fertility” section, show a much lower number of invalid skips in all parts except in the abortion questions. While invalid skips do not reach the level seen in table 5.4.1, on average 190 female respondents were not asked each abortion question (190 is an average from all abortion questions, not just those shown in the tables). The table also shows a number of other trends. First, respondents have higher levels of don’t know answers the more precise the question being asked. For example, in Table 5.4.2, when males were asked the date of birth of their first child, only one did not know the year, three did not know the month and 10 did not know the day. This phenomena is most clearly seen in table 5.4.5, which shows the year and month of the respondent’s first sexual encounter. Only 43 respondents did not know the year, but 1,410 respondents did not know the month. This problem with dates is also seen in the abortion data where only four respondents did not know the year when they had their first abortion, but 13 did not know the month.

Refusal rates in the “Fertility” section are quite low except for a number of key questions. Asking the number of times they had sex in the last month elicited high rates of refusal for males and females. This question elicited 167 male and 135 female refusals. Interestingly, most individuals were willing to answer if they ever had sex since only 45 males and 54 females refused to answer these questions. Birth control questions did not have exceptionally high rates of refusal. Seventeen female respondents and no males refused to answer the birth control questions. Table 5.4.6 shows that 28 females refused to answer if they ever had an abortion and 28 more refused to state if they dropped out of school before they terminated the pregnancy.

Table 5.4.2 Amount of Missing Data Per Question in Male Fertility Section, 1984 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R13017.	Ever Had Any Children	0	3	0
R13019.	Month Birth Child#1 Born	41	3	0
R13020.	Day Birth Child#1 Born	45	10	0
R13021.	Year Birth Child#1 Born	39	1	0
R13022.	Sex of Child#1 Born	3	0	0
R13115.	Total #Children Expect to Have	12	45	3
R13117.	#Years Expect Have 1 st /Next Child	22	120	0
R13118.	Had Any Children/Expecting	0	7	0
R13119.	Current Pregnancy Planned	131	0	0
R13121.	Ever Had Sexual Intercourse	12	0	45
R13122.	Age @First Sexual Intercourse	28	19	23
R13123.	#Times Sexual Intercourse Past Month	11	68	167
R13124.	Is Partner Now Pregnant	0	1	0
R13125.	Use Any Birth Control During Last Month	15	2	0
R13126.	#Times Try Prevent Pregnancy	65	0	0
R13127.–R13141.	Method of Birth Control	16	0	0
R13142.	Ever Have a Sex Education Course	10	0	12
R13148.	Month Took Sex-Ed Course	73	564	0
R13149.	Year Took Sex-Ed Course	36	58	0
R13150.	Time When Pregnancy Most Likely	19	1480	20

Table 5.4.3 Amount of Missing Data Per Question in Female Fertility Section, 1984 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R13191.	#Pregnancies	8	0	0
R13251.	Use Any Birth Control before Preg#1	18	0	1
R13254.	Want Be Pregnant before Preg#1	20	0	0
R13255.	Husband/Partner Want Preg#1	19	20	0
R13283.	Get Prenatal Care Preg#1	57	0	0
R13286.	Frequency Alcohol Use Preg#1	58	0	0
R13288.	#Cigarettes Smoked Preg#1	56	0	0
R13297.	X-Rays Taken Preg#1	57	0	0
R13302.	Sonogram Preg#1	57	6	0
R13358.	Amniocentesis Preg#1	57	0	0
R13411.	Took Vitamins Preg#1	57	0	0
R13443.	C-Section Child#1 Born	52	0	0
R13445.	Weight at Delivery, Preg#1	53	5	1
R13446.	Weight before Preg#1	51	5	1
R13449.	Length Child#1 Born at Birth	53	20	0
R13667.	Weight of Child#1 @Birth Lbs	25	6	0

Table 5.4.4 Amount of Missing Data Per Question in Feeding Part of Fertility Section, 1984 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R13670.	Child#1 Breastfed	27	0	0
R13672.	Month Age Child#1 Breast Fed Ended	27	1	0
R13674.	Month Age Child#1 Formula Fed	38	3	0
R13693.	Wk Age Child#1 Formula Fed Ended	57	0	0
R13694.	Month Age Child#1 Formula Fed Ended	57	6	0
R13696.	Months Age Child#1 – Cow's Milk	81	10	0
R13698.	Months Age Child#1 – Solid Food	86	10	0

Table 5.4.5 Amount of Missing Data Per Question in Child Part of Fertility Section, 1984 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R13791.	Age Had 1 st Menstrual Period	8	14	22
R13792.	Year 1 st Menstrual Period	0	7	0
R13793.	Month Had 1 st Menstrual Period	17	2207	1
R13794.	R Ever Been Pregnant	0	1	0
R13795.	Ever Had Sexual Intercourse	4	0	54
R13796.	Age First Sexual Intercourse	5	26	78
R13797.	Year 1 st Sexual Intercourse	0	43	66
R13798.	Month Sexual Intercourse 1 st Time	19	1410	75
R13799.	#Times Sexual Intercourse Past Month	9	104	135
R13802.	#Times Try Prevent Pregnant Past Month	17	0	2

Table 5.4.6 Amount of Missing Data Per Question in Abortion Questions of Fertility Section, 1984 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R13827.	Ever Had An Abortion	135	0	28
R13828.	# of Abortions	143	0	0
R13830.	Year of 1 st Reported Abortion	196	4	0
R13837.	Drop out School #1 Pregnant	155	0	28
R13839.	Year Left School 1 st Time Pregnant	164	0	0
R13841.	Year Return School Time#1 after Pregnant	258	0	0

Tables 5.4.7 and 5.4.8 examine the “Income & Assets” section of the 1989 survey. While invalid skips are relatively rare in this section, refusals and don’t know answers are fairly prevalent. The question with the highest amount of missing income data is R29822., which asks how much income was earned by other adults living in the household who were related to the respondent. While the previous questions showed that most respondents knew the type of income received by these family members, 958 could not come up with a specific amount. The second most problematic question with 11 invalid skips, 155

don't knows, and 113 refusals was R29714., which asked the respondent how much they earned from wages, salary, and tips. Other questions with high numbers of don't knows are R29813., which asked about the amount of money received from other sources like interest and dividends, R29825., which asks about a partner's income, and R29827., which asks the number of exemptions used when filing a federal tax return.

The asset table (Table 5.4.8) also shows invalid skips are rare but don't know and refusal rates are not. Surprisingly, one of the questions with the highest amount of missing data (315 missing answers) asks, "how much is your car worth (R29852.)?" Another question missing many observations asks the amount of the respondent's savings (R29835.). While the car worth question primarily elicits don't knows, the savings question resulted in 160 refusals. Three other questions elicited high numbers of don't knows: value of stocks and bonds (R29837.) – 219 don't knows; how much money taken out of savings last year (R29842.) – 222 don't knows; and the market value of other items such as jewelry (R29854.) – 151 don't knows.

Table 5.4.7 Amount of Missing Data Per Question in Income Section, 1989 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R29714.	Amount Rec from Wages/Salary/Tips	11	155	113
R29715.	In 1988 Receive Income from Own Business	1	0	11
R29717.	How Much Did R Receive after Expenses	6	49	23
R29732.	Amount Rec'd Per Week from Unemployment	0	5	1
R29736.	Amount Sp Rec'd 1988 from Wages	16	17	70
R29754.	How Much Did Sp Receive from Unemployment	8	12	0
R29758.	R/Spouse Rec'd Money for Child Support	1	1	10
R29759.	Amount R/Spouse Rec'd Child Support	2	14	2
R29760.	R/Spouse Rec'd AFDC Payments	0	4	9
R29774.	R/Spouse Rec'd Food Stamps	0	2	10
R29788.	R/Spouse Rec'd SSI/Public Assistance	0	4	9
R29808.	Rec'd Veteran Benefits	1	1	10
R29812.	R/Spouse Rec'd Money from Oth So	0	2	16
R29822.	Income Rec'd by Adults Related T	7	958	8
R29825.	Total Income Rec'd before Deduct	2	200	4
R29826.	Sp File Federal Income Tax R	0	2	13
R29827.	R'S Filing Status on Federal Ret	11	8	2
R29828.	Exemptions Filed on 1988 Federal Tax	62	92	3

Table 5.4.8 Amount of Missing Data Per Question in Asset Section, 1989 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R29831.	Amount Property Selling for on Today	5	53	10
R29832.	Amount R Owes on Property	4	85	25
R29833.	Amount Other Debt R Owes on Property	12	26	27
R29835.	Amount of Savings	7	166	160
R29837.	Current Market Value of Stocks	2	219	23
R29838.	R/Spouse Have Rights to Estate	2	3	18
R29839.	Total Value of Estate	3	90	6
R29840.	Put Money in/out of Savings	1	3	28
R29841.	How Much More Money Put in	6	110	53
R29842.	How Much More Money Take out	5	222	21
R29843.	R Have Business Investment	0	1	12
R29844.	R Have Investment in a Farm	4	0	0
R29847.	Total Market Value of Business	4	75	10
R29848.	Total Amount of Business Debt	1	55	8
R29851.	How Much Does R Owe on Vehicle	0	56	17
R29852.	Amount Vehicle Sells for Today	11	293	11
R29854.	Market Value of Other Items	5	151	25
R29856.	Total Amount R Owes	1	73	13

Table 5.4.9 and 5.4.10 examine the drug and alcohol use supplements in the 1994 survey. In these CAPI modules, there are no invalid skips. Interestingly, there are extremely low refusal and don't know rates within the "Alcohol" section (Table 5.4.9). The question with the highest refusals (nine respondents) asks if the individual had a drink since the 1989 interview. The typical question in the "Alcohol" section received only two refusals. Don't know rates are also low. The maximum number of don't knows at nine occurs in R49803., which asks if the respondent needs to drink more alcohol now in order to get drunk. On average, the "Alcohol" section records only 1.5 don't knows per question.

These low numbers of refusals and don't knows are not seen in Table 5.4.10, which examines the "Drug Use" section. On average, the typical question in this supplement elicited 23 don't knows and 48 refusals. Readers should understand that this supplement was filled in directly by the respondent, not by the interviewer. To provide respondents with practice using a CAPI computer, the questionnaire asked them two non-drug use practice questions. Refusal rates are even high for these two test questions, which ask how many more children the respondent expects to have and what type of entertainment, such as movies, concerts, or plays, the respondent went to last year.

The highest number of refusals (119) occurs in R50532., which asks the age the respondent first used marijuana. The second largest number of refusals occurs in a similar question, R50536., which asks the

age of first cocaine use. These same questions have very high don't know responses (113 marijuana and 48 cocaine). One other question with a very high don't know rate is R50525., which asks if the respondent ever smoked cigarettes daily. Almost 80 individuals did not know the answer to this question. Given that the question wording is straightforward, it is likely a number of respondents are using don't know as a polite way of refusing to answer the question.

Table 5.4.9 Amount of Missing Data Per Question in Alcohol Use Section, 1994 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R49791.	R Had Drink of Alcohol since 1989	0	3	9
R49792.	Had Alcoholic Beverage in Last 30	0	0	5
R49793.	Times Had 6/More Drinks Last	0	0	1
R49794.	How Many af Last 30 Days Drank A	0	6	2
R49795.	No. of Drinks on Avg. Day When R	0	8	3
R49803.	Need More to Get Drunk Than Before	0	9	0
R49808.	Arrested, in Police Trouble	0	0	3
R49809.	Drink More Than Before	0	4	3

Table 5.4.10 Amount of Missing Data Per Question in Drug Use Section, 1994 Survey

Reference #	Variable Title	Invalid Skip	Don't Know	Refusal
R50524.	R Smoked at Least 100 Cigrts in Life?	0	24	38
R50525.	R Ever Smoked Daily?	0	79	49
R50526.	Age When R 1 st Started Smoking Daily?	0	33	12
R50531.	Total Occasion R Use Marijuana	0	33	89
R50532.	Age 1 st Time Used Marijuana	0	113	119
R50533.	Most Recent Time Used Marijuana	0	35	89
R50535.	How Many Occasions Used Cocaine	0	19	86
R50536.	Age 1 st Time Used Cocaine	0	48	103
R50537.	Most Recent Time Used Cocaine	0	15	78
R50539.	How Many Occasions Used Crack	0	15	77
R50540.	Age 1 st Time Used Crack	0	33	82
R50541.	Most Recent Time Used Crack	0	16	74
R50553.	R Used Heroin w/o Doctor's Instr	0	9	53

References

- Mott, Frank L. "Patterning of Child Assessment Completion Rates in the NLSY: 1986–1996." CHRR, The Ohio State University, 1998.
- Mott, Frank L. "Evaluation of Fertility Data and Preliminary Analytical Results from the 1983 (5th round) Survey of the National Longitudinal Survey of Work Experience of Youth." CHRR, The Ohio State University, 1985.
- Mott, Frank L. "The Patterning of Female Teenage Sexual Behaviors and Attitudes." CHRR, The Ohio State University, 1994.
- Mott, Frank L. "Fertility-Related Data in the 1982 National Longitudinal Surveys of Work Experience of Youth: An Evaluation of Data Quality and Some Preliminary Analytical Results." CHRR, The Ohio State University, 1983.
- Olsen, Randall J. "The Effects of Computer Assisted Interviewing on Data Quality." CHRR, The Ohio State University, 1992.